Document statistics: 338 lines 8,599 words

Running head: PERCEPTION OF THE VISUAL ENVIRONMENT

Perception of the visual environment

Benjamin W. Tatler

School of Psychology

University of Dundee, UK

Abstract

The eyes are the front end to the vast majority of the human behavioural repertoire. The manner in which our eyes sample the environment places fundamental constraints upon the information that is available for subsequent processing in the brain: the small window of clear vision at the centre of gaze can only be directed at an average of about three locations in the environment in every second. We are largely unaware of these continual movements, making eye movements a valuable objective measure that can provide a window into the cognitive processes underlying many of our behaviours. The valuable resource of high quality vision must be allocated with care in order to provide the right information at the right time for the behaviours we engage in. However, the mechanisms that underlie the decisions about where and when to move the eyes remain to be fully understood. In this chapter I consider what has been learnt about targeting the eyes in a range of different experimental paradigms, from simple stimuli arrays of only a few isolated targets, to complex arrays and photographs of real environments, and finally to natural task settings. Much has been learnt about how we view photographs, and current models incorporate low-level image salience, motor biases to favour certain ways of moving the eyes, higher-level expectations of what objects look like and expectations about where we will find objects in a scene. Finally in this chapter I will consider the fate of information that has received overt visual attention. While much of the detailed information from what we look at is lost, some remains, yet our understanding of what we retain and the factors that govern what is remembered and what is forgotten are not well understood. It appears that our expectations about what we will need to know later in the task are important in determining what we represent and retain in visual memory, and that our representations are shaped by the interactions that we engage in with objects.

Perception of the visual environment

The human behavioural repertoire is intricately linked to the gaze control system: many behaviours require visual information at some point in their planning or execution. The information that we require for successful completion of behavioural goals is likely to be drawn from two sources: visual information available on the retina for the current fixation, and information stored from previous fixations. Thus in order to understand how information is gathered from the environment, we must understand both how gaze is allocated in order to sample information, and the fate of information once sampled but no longer fixated.

In the sections that follow I will first consider how information is sampled from the visual environment. In particular the mechanisms that might underlie targeting decisions for gaze allocation will be discussed. Following this, the fate of information sampled in each fixation will be considered. In particular I will discuss how information is encoded into memory and retained as representations of the objects and environment.

Sampling information from the visual environment

The visual information supplied by the eyes is limited both in space and time. While some tasks can be carried out effectively in peripheral vision, such as maintaining heading using lane edges when driving (Land & Horwood, 1995), any tasks that require finely detailed information necessitate that the high acuity fovea must be directed toward locations that contain this behaviourally relevant information. Not only is high quality visual information sampling restricted in space to the central foveal region of the retina but it is also restricted in time. For useful visual information to be gathered, the image on the retina must be kept relatively stable: we see little or nothing when the eyes are moving (Erdmann & Dodge, 1898). To balance this need to keep the eyes still in order to gather information, with the need to move the foveae to the areas of the environment from

which detailed information is required, foveal vision is typically directed to around 3-4 locations in every second, with fixation pauses between these movements lasting for an average of around 200-400 ms (Rayner, 1998; Land & Tatler, 2009). These strict spatial and temporal limits on sampling place a clear emphasis upon the need for effective allocation of the valuable resource of high quality vision.

In this chapter when discussing visual sampling from the environment I will primarily discuss how central, foveal vision is allocated and used to gather information. This is not to devalue the role of peripheral vision or to suggest that locations outside the fovea are unprocessed or unencoded. However, understanding where we point our foveae is important not only for tasks that require finely detailed information, but for many of the behaviours that we engage in. This is because, even when peripheral vision is sufficient to extract information from a location, we tend to point our foveae at things that we are manipulating or require information from (Ballard, Hayhoe, Li, & Whitehead, 1992). There are a variety of reasons for this. For example, Ballard et al. (1992) found that when moving blocks on a screen to copy a pattern, participants could complete the task using peripheral vision, but took longer to complete the task than if they were allowed to foveate the blocks that they were manipulating. When driving, people look at the tangent point to the bend as they approach it (Land & Lee, 1994), not because there is finely detailed information there that they need, but because the angle between the car's current heading and the tangent point directly informs how much the steering wheel should be rotated to steer around the bend correctly. So the angle between the driver's body orientation and their gaze direction provides the information needed to steer. Whatever, the reason for foveating a location, the intimate link between where we look and what we do (Ballard et al., 1992; Land & Tatler, 2009) places particular importance upon understanding what factors underlie decisions about where to point the eyes.

The importance of characterising the allocation of foveal vision in space and time

has been recognised since the saccade and fixate strategy of the eye was first characterised objectively in the late 19th Century (Hering, 1879; Wade & Tatler, 2005). When viewing complex scenes such as photographs, we see that fixation allocation is far from random. Within a single participant, viewing patterns are very similar when viewing the same scene several times, (Yarbus, 1967), suggesting common fixation selection criteria on multiple viewings of a scene. Similarly, fixation distributions for multiple participants show overall similarity: when a number of participants each view the same scene, they will tend to select similar locations to fixate (Buswell, 1935; Yarbus, 1967). Such between-observer consistency in viewing behaviour suggests common underlying principles for selecting where to fixate in complex scenes. The question of what these common underlying principles might be has been the focus of considerable research effort over the past few decades, and has given rise to a number of computation models of fixation selection.

Paradigmatic considerations

Before discussing what we currently understand about visual sampling strategies when viewing scenes, it is important to consider what we mean by a scene and what we want to understand about visual sampling. When people talk about natural scenes or real world scenes, this can mean static photographic scenes, dynamic movie sequences, or real three-dimensional environments. The differences between these three classes of scene are immense.

Physically, static photographic scenes necessarily lack binocular depth and motion cues and occupy a much narrower dynamic range than real environments. Dynamic scenes have the advantage of providing motion cues but these may be rather different from those experienced in real environments. Compositional biases abound in photographic (Tatler, Baddeley, & Gilchrist, 2005) and dynamic (Dorr, Martinetz, Gegenfurtner, & Barth, 2010) scenes, whereby there is a greater prevalence of low level featural information in the centre

of the scene than in the periphery. Both static and dynamic scenes artificially control the observer's viewpoint of the scene and limit the visual environment to the frame of the monitor in which the image or movie is displayed.

In addition to these physical differences, static and dynamic scenes are presented using paradigms that include events which do not occur in natural environments. Specifically, static scene paradigms typically involve the sudden onset of a scene followed by inspection and then removal of the scene a few seconds later. Dynamic scenes typically involve sudden onsets at the start, but may also contain frequent editorial cuts which abruptly change the viewpoint of the observer. Neither sudden onsets, nor abrupt viewpoint changes are a feature of viewing natural environments.

The task of the observer is often rather different across these three classes of scene. In a natural environment we typically employ gaze to aid our motor actions and allow us to achieve defined behavioural goals (Land & Tatler, 2009). Such physical interaction is necessarily absent in most paradigms that involve static photographs or dynamic movies as stimuli. The lack of motor interaction with the scene may well have fundamental effects upon the behaviour of the gaze system (Steinman, 2003).

Of course, it is not the case that we can simply and strongly differentiate these three categories of scene, nor claim that scenes on screens (whether static or dynamic) are not components of our everyday behaviours and environments. For work and entertainment purposes we view content on screens for much of the time and many of the environments that we find ourselves in during everyday life contain screens. Indeed there are tasks that we perform that rely on screen-based viewing like CCTV surveillance (Stainer, Scott-Brown, & Tatler, 2013). Thus screen-based viewing paradigms can be informative of certain everyday behaviours. The important point is to remember the scope of the paradigm being used. The differences in the physical characteristics, compositional biases, protocols and goals when viewing static images, dynamic movies and real environments

mean that it is not clear to what extent findings for one type of scene can be generalised to the others. Thus, it is important to consider evidence from paradigms that are appropriate for the domain of explanation that one is interested in.

In the sections that follow I will discuss in turn what is currently understood about how we sample information from static scenes, dynamic scenes and real environments (any environment that extends beyond the limits of a single screen). Most of the models that have been proposed for how we direct our eyes around scenes have been derived from data collected using static scene viewing paradigms.

Static scenes

Static images provided the first insights into how we look at complex scenes. In a landmark series of studies, Buswell (1935) recorded eye movements as people viewed a series of photographic and painted scenes(e.g. see Figure 1).

Buswell's work provided many insights about how people view complex scenes, many of which echo themes present in current scene perception research (see Wade & Tatler, 2005, for discussion of these various contributions). One important insight was to recognise that certain regions in scenes are fixated by most participants, which Buswell described as *Centers of Interest*. Buswell used patterns, such as geometric motifs in architecture to consider whether there was anything distinctive about the visual motifs that attracted fixation. Buswell's conclusions on this matter were mixed: in some cases he felt there was a clear link between the lines and motifs in a pattern and where people looked, but in other cases he felt the link was much weaker than he had expected. This consideration of the link between visual motifs and fixation patterns shows that the question of the extent to which fixation behaviour is driven by low- or high-level factors has been present since eye movements were first recorded when viewing complex scenes.

The extent to which eye movements are driven by low-level visual information or

higher-level factors continues to be a central theme in modern eye movement research (see Tatler, Hayhoe, Land, & Ballard, 2011). While no-one would argue either extreme position, the relative contributions of low- and high-level factors remains the subject of considerable debate and controversy (see Tatler, 2009). A particular challenge in the field has been to construct computational models that account for human fixation behaviour while viewing static scenes.

The majority of existing models of fixation selection are based around (but not restricted to) the notion that low-level feature information in scenes has an important influence on fixation selection (for reviews of state of the art models see Borji & Itti, 2013; T. Judd, Durand, & Torralba, 2012). In general, these models are based around the idea that what attracts the eye is any location that stands out visually from its surroundings; i.e. a location that is visually conspicuous. I will refer to this class of models as conspicuity-based models. The most prominent of these is Itti and Koch's visual salience model (Koch & Ullman, 1985; Itti, Koch, & Niebur, 1998; Itti & Koch, 2000), which is in many ways the precursor to most current models. In this model, the low-level information in a scene is operationalised via a set of biologically-plausible filters that extract local luminance-, colour- and orientation-contrast in the scene. Feature maps are combined across features and spatial scales via local competition in order to produce a single overall visual conspicuity map referred to as a *salience map* (see Figure 2). Allocation of attention (either overt or covert) then proceeds from this salience map using a winner-takes-all process: attention is allocated to the spatial location in a scene corresponding to the maximum peak in the salience map. A local inhibition of return mechanism then suppresses activity in the salience map at attended locations, resulting in a relocation of attention to the next most salient location, and so on. The model therefore proposes that attention is allocated to locations in a scene on the basis of visual conspicuity and in order from the most salient location in a scene to the least. Covert

attention and where we look are typically assumed to be intricately linked, with attention allocated covertly to locations just prior to directing the fovea to the attended location (Deubel & Schneider, 1996). In this way, the salience model functions equally as a model of covert attention allocation or overt allocation of attention (i.e. where we look).

The salience model is an attractive account of human fixation selection for at least three reasons. First, the model is biologically plausible in that the kind of low-level feature extraction operationalised in the model is rather similar to the kinds of features that we know the early visual system can extract. Second, the model offers a logical extension to the results found for the principles that might underlie attention allocation in more simple search conditions. In simple search arrays where the target differs from the distractors in a single feature dimension, there is clear evidence for pre-attentional capture ("pop-out") by low-level information (Treisman & Gelade, 1980). When searching for targets defined by the unique conjunction of two features, search is harder and requires multiple relocations of attention (Treisman & Gelade, 1980). However, for both pop-out and feature conjunction search, models based purely on low-level feature information have been successful (Treisman & Gelade, 1980; Wolfe, 2007). It is a natural extension of this work to suggest that the same low-level principles that underlie models such as Wolfe's guided search model (2007) for search arrays might also underlie fixation selection in more complex scenes. Third, the model offers a computable solution to describing properties of scenes. That is, low-level image features are computable and local conspicuity in scenes can be quantified. In contrast, higher-level understanding of scenes and behavioural goals are hard to quantify or describe computationally.

The salience model has been used to successfully detect pop-out visual search targets in a single iteration of the winner-takes-all process, and to replicate multi-fixation search patterns for more complex search targets (Itti & Koch, 2000). The impact of the salience model both within and outside the context of vision research has been extensive.

The salience model has been used as an automated system to find military vehicles in complex scenes (Itti & Koch, 2000). The principles of the salience model have been applied to robotic visual systems (Frintrop, Rome, & Christensen, n.d.; Siagian & Itti, 2007; Xu, Kuehnlenz, & Buss, 2010) and used in medical applications to locate tumours in scans (Hong & Brady, 2003).

Evaluations of the salience model (and similar models based on low-level feature-based fixation selection) in complex scenes have been prevalent in recent literature (see Tatler et al., 2011). Typically, the explanatory power of such models is evaluated using one of two methods: measuring local image statistics at fixation (Reinagel & Zador, 1999); or using the model to predict where humans should fixate and seeing how well human fixation behaviour matches these predictions (Torralba, Oliva, Castelhano, & Henderson, 2006). In both cases evidence can be found that seems to support the notion that low-level information has a role to play in fixation selection. Fixated locations tend to have higher salience (greater visual conspicuity) than control locations(Parkhurst, Law, & Niebur, 2002), and more fixations tend to be made in locations predicted by conspicuity models than would be expected by chance (Foulsham & Underwood, 2008).

Despite these attractions and successes of conspicuity-based models such as the salience model, these results must be interpreted with caution. First, the explanatory power of such models is relatively weak. If the magnitude of the differences in image statistics between fixated and control locations is considered it becomes clear that the differences are quite small (Einhauser, Spain, & Perona, 2008; Nyström & Holmqvist, 2008; Tatler, Baddeley, & Gilchrist, 2005). Many existing conspicuity-based models, are no better able to describe human fixation behaviour than a Gaussian centred on the middle of the scene (see Bylinskii et al., n.d.); thus, knowing that people look in the middle of the screen (Tatler, 2007) explains more fixation behaviour than most contemporary computational models. In a recent evaluation of conspicuity-based models

using a database of 2000 images (Bylinskii et al., n.d.), only one model outperformed a central Gaussian (T. Judd, Ehinger, Durand, & Torralba, 2009) and did so by a very small margin (a central Gaussian accounted for human fixations with an AUC of 0.83, whereas the Judd et al. model classified with an AUC of 0.84). Second, the interpretation of the basic findings is problematic: correlations between low-level information and fixation selection need not imply causal links (Henderson, Brockmole, Castelhano, & Mack, 2007; Henderson, 2003; Tatler, 2007) but may arise due to the correlations that exist between low-level features and higher-level scene content. Indeed maps of where objects are in scenes accounts for more human fixations than maps of low-level conspicuity in the same scenes (Einhauser et al., 2008), and fixations tend to target the centres of objects, suggesting an important role for object-level information in saccade target selection (Nuthmann & Henderson, 2010). Moreover, conspicuity-based models fail to account for how fixation selection changes with changes to the observer's goals when viewing the scene (Foulsham & Underwood, 2008; Henderson et al., 2007).

Not only is the explanatory power of visual conspicuity models limited, but the models often contain a set of problematic assumptions that do not hold up to empirical or theoretical scrutiny (Tatler et al., 2011). For example many models fail to account for limited peripheral acuity when computing salience maps (see Wischnewski, Belardinelli, & Schneider, 2010, for discussion of this issue), and neglect issues such as time, order and spatial precision of fixation selection (see Tatler et al., 2011). The inclusion of inhibition of return is necessary for computational models based on winner-takes-all selection, yet there is no compelling evidence that humans show any decreased tendency to re-fixate a recently-fixated location when viewing complex scenes (Hooge, Over, Van Wezel, & Frens, 2005; Smith & Henderson, 2009; Tatler & Vincent, 2008). Perhaps the key theoretical assumption is that models should be built around a core selection principle based on low-level feature information. Given the empirical shortcomings described above, there is

little evidence for any substantial role of low-level features in driving fixation selection. It therefore seems somewhat surprising that models have retained such a prominent role for low-level features.

Task effects on static scene viewing

The importance of the observer's behavioural goals when viewing an image has been recognised since Buswell's seminal work. Buswell (1935) showed that when an individual views the same scene, but with different instructions, the inspection patterns are very different (see Figure 3).

The impact of task instructions on fixation behaviour became even more apparent when Yarbus (1967) conducted a similar experiment in which a participant viewed the same painted scene seven times, each time with a different instruction prior to viewing (Figure 4). From this elegant demonstration it was clear that behavioural goals have a dramatic influence on viewing behaviour.

The fundamental limit of stimulus-driven models of fixation behaviour is that they cannot readily account for the differences evident in figures 3 and 4, that arise due to variations in task instructions. This was recognised from the outset (Itti & Koch, 2000), and has underpinned the development of new models of fixation selection that attempt to account for high-level effects such as task.

Modelling high-level effects in static scene viewing

Several models have proposed ways of incorporating high-level factors into models of eye movement behaviour when viewing scenes. Navalpakkam and Itti (2005) proposed that high-level effects may be manifest as differential weightings of the individual feature channels that combine to produce the salience map. If the features of a target are known, this knowledge can be used to weight relevant features; this should enhance the representation of the object in the resultant salience map. Torralba and colleagues (2006)

suggested that the visual system may exploit the typical spatial relationships that exist between objects and the scenes in which they occur. Most objects are not equally likely to occur in all scene regions: they will be very unlikely to be in certain locations and very likely to be in others. For example, a clock is far more likely to be found on a wall than on the floor or ceiling. Torralba and colleagues suggested that learnt associations between objects and spatial regions of scenes are used to "narrow down the search" for an item. Computationally, this is operationalised as a spatial mask corresponding to the likely scene region, which is then used as a modifier for the overall salience map, such that the gaze system then targets salient locations that occur within the scene regions that are likely to contain the target object. This class of model is able to produce very good descriptions of fixation behaviour when searching for objects in scenes, particularly in the first few fixations of a viewing epoch (Torralba et al., 2006); for example, when searching for a painting in a scene, the first fixation of the search process was accounted for in just under 40% of cases by salience alone, but in just over 70% by a model comprising salience and expected target location (Torralba et al., 2006). Cottrell and colleagues also proposed a scheme in which prior knowledge is used to guide fixation selection (Kanan, Tong, Zhang, & Cottrell, 2009). However, they proposed that spatial expectancy is not the only useful source of knowledge when searching for an object: prior knowledge of objects of the same class can be used to provide a template for search based on the expected characteristics of the target object. Again, the resultant object appearance map is operationalised as a spatial mask, which is used to modify a salience-like map of the entire scene, so that fixations target locations that contain salient low-level information but are also within regions identified as sharing characteristic properties with the target object class. Kanan *et al.*'s (2009) model based on these principles again offered the ability to account for an impressive proportion of human fixations generated when searching for targets in photographic scenes, with salience alone accounting for around 55% of human

fixations when searching photographic scenes, but just over 70% of fixations being accounted for by a model comprising salience, expected location and expected appearance of targets. Ehinger et al. (2009) combined salience, target appearance and expected target location in a single model and were able to account for a large fraction of human fixation behaviour when searching scenes for people: their average AUC was 0.90, which roughly equates to an ability to account for around 90% of human fixations during this task.

While the models discussed above all retain salience at their core, an alternative approach to modelling fixation behaviour can be found in Zelinsky's (2008) Target Acquisition Model. This model accounts for retinal inhomogeneity of sampling and includes high-level knowledge about the target of a search. The departure from the above models is that visual information is not represented as simple feature maps. Rather, higher-order derivatives are represented which incorporate object knowledge. This selection is based on much higher-level representations than the modified salience map in the scheme described above. This model has been successful at replicating a number of aspects of human fixation behaviour across a range of visual stimuli.

Problems with models of static scene viewing

While models of scene viewing such as those discussed above are able to account for a reasonable fraction of eye movement behaviour, it is worth returning to the issues that arise with the static scene viewing paradigm. In particular, I will consider the problems associated with sudden onsets and with the framing effect of the monitor in which the images are displayed.

Viewing behaviour soon after a sudden onset is different from that observed later in a viewing period. This was first demonstrated by Buswell (1935, see Figure 5) who showed that there was a higher degree of consistency between subjects in where they chose to fixate early in viewing than there was later in viewing. This early between-subject

consistency in fixation placement followed by later divergence between subjects has been found repeatedly in more recent studies (e.g. Tatler, Baddeley, & Gilchrist, 2005, Figure 5). Why these differences exist has been the topic of some debate and controversy. One possibility is that the relative contributions of low- and high-level factors in saccade targeting changes over time, such that fixations soon after scene onset are driven more by low-level factors whereas later fixations are driven more by high-level factors (Carmi & Itti, 2006; Parkhurst et al., 2002). However, evidence to the contrary also exists (Tatler, Baddeley, & Gilchrist, 2005; Nyström & Holmqvist, 2008). These authors suggest that there is no evidence for any change in the contribution of low-level factors over viewing time; the observed changes from initial between-observer consistency in fixation to later inconsistency must emerge due to higher-level factors such as *strategic divergence* in fixation selection processes (Tatler, Baddeley, & Gilchrist, 2005).

While there is disagreement about the source of changes in viewing behaviour over time, there is agreement that that viewing behaviour soon after the onset of a scene is different from that recorded several seconds later. If viewing behaviour soon after an onset is unlike that during extended viewing, then the presentation times of scenes may have a substantial influence on the fixation behaviour recorded in experiments involving images. With very short presentation times, common targeting mechanisms are likely to be found across participants. With longer presentation times, more divergence in fixation selection will be observed, presumably reflecting greater divergence in targeting principles. However, this should not be taken as a recommendation for short presentation times in static scene viewing paradigms. One way to interpret the differences between early and late inspection behaviour for scenes is to suggest that the early apparent consistency between observers is driven by the onset rather than the stimulus or task *per se*. As such, the behaviour recorded immediately after scene onset may not be representative of the normal mechanisms that underlie inspection behaviour.

A second issue associated with static scene viewing is that the images are almost always displayed within the bounds of the (usually visible) monitor frame. This frame in itself appears to have quite an influence on fixation behaviour when viewing scenes. Many authors have reported that distributions of fixations during scene viewing show considerable spatial biases toward the centre of the scene (e.g. Parkhurst et al., 2002). However, the reasons for this spatial bias were initially unclear because photographic scenes typically show compositional biases (Tatler, Baddeley, & Gilchrist, 2005). These compositional biases arise form the natural tendency to place objects of interest near the centre of the viewfinder when taking photographs. The result is that the spatial distribution of low-level information in scenes tends to show a central weighting. It is therefore unclear whether the central bias in fixation behaviour is associated with the central bias in visual content. To address this issue, Tatler (2007) used scenes with unusual spatial biases in their feature content. Tatler (2007) showed that irrespective of the feature biases in the scenes, the observers showed the same strong bias to fixate the centre of the scene (Figure 6). This result not only highlights the lack of correlation between low-level image features in scenes, but also suggests that a significant proportion of fixation behaviour recorded when viewing scenes may be driven by the monitor frame (or expected composition of the scene) rather than the content of the scene. If fixation distributions contain biases arising from factors related to the framing and expected composition of a photograph, caution is required when interpreting data derived from static scene viewing and when designing the layout of experimental materials.

Dynamic scenes

Given the concerns raised about static scenes as stimuli for eye movement experiments, dynamic scenes are increasingly being used as alternative stimuli for investigating how we view scenes. Dynamic features can be considered as an additional

low-level feature in computational models of fixation selection, and under some circumstances the addition of dynamic features to the salience model can improve its ability to account for human viewing behaviour (Itti, 2005). However, the situations in which dynamic features add explanatory power to computational models are those in which there are frequent editorial cuts that instantaneously change the observer's viewpoint. Such editorial cuts themselves introduce artefacts into the eye movement record in a manner not dissimilar to those arising from the sudden onset of static scenes. Movies with cuts tend to produce eye movement behaviour with strong central fixation biases (Dorr et al., 2010; t Hart et al., 2009). Moreover, eye movement behaviour when viewing movies with editorial cuts is not like that produced when viewing movies shot from a single viewpoint, without any editorial cuts (Dorr et al., 2010). Furthermore, unlike when viewing edited movies, when viewing continuous movies, dynamic features are not predictive of fixation behaviour (Cristino & Baddeley, 2009). Here the strongest predictors of fixation behaviour are the screen centre and a spatial bias related to the perceived horizon in the scene (Cristino & Baddeley, 2009).

Modelling eye movement when viewing dynamic scenes

The latest models of fixation selection attempt to explain eye movements made to dynamic scenes (e.g. Wischnewski et al., 2010; Wischnewski, Steil, Kehrer, & Schneider, 2009). Like Zelinsky (2008), Wischnewski et al. depart from first order features as the domain for targeting decisions. Rather, they propose that targeting is based upon a representation comprising second (or higher) order static and dynamic features, combined with top down task information. The resulting attentional priority map is conceptually similar to that described by Fecteau and Munoz (2006). This model has demonstrated impressive ability to account for human fixation behaviour while viewing dynamic scenes and is a promising direction for such models.

Natural behaviour

If we wish to understand the manner in which gaze is employed to aid our activities in real situations, we must consider eye movement behaviour in natural, everyday settings. The paradigmatic limitations of static images and even dynamic movies are such that it is unclear whether findings from these paradigms will generalise to behaviour conducted in environments that extend beyond the limits of a single screen. Certainly, eye movement behaviour observed when interacting with objects is fundamentally different from that observed when simply inspecting the same objects (Epelboim et al., 1997, 1995), suggesting that acting upon objects changes how we insect the environment. There is growing interest in studying eye movements in the context of everyday behaviour (Land & Tatler, 2009) and we are now in a position to consider whether the principles that seem to underlie fixation selection in static and dynamic screen-based scene viewing paradigms are consistent with eye movement behaviour in real-world settings. What is clear across a range of everyday tasks is that there is close spatial and temporal coupling between vision and action: we tend to look at the object we are manipulating (e.g. Ballard et al., 1992; Land & Furneaux, 1997; Land, Mennie, & Rusted, 1999; Patla & Vickers, 1997; Hayhoe, Shrivastava, Mruczek, & Pelz, 2003; Pelz & Canosa, 2001). The link between behavioural goals and fixation placement is very strong: in everyday activities essentially all fixations target task-relevant objects in the environment (Hayhoe et al., 2003; Land et al., 1999). Moreover, placement of fixations within an object depend upon the intended purpose of interaction with that object. For two classes of visually similar objects Rothkopf *et al.* (C. A. Rothkopf, Ballard, & Hayhoe, 2007) showed that fixations were directed to the margins of objects that the observer intended to avoid, but to the centres of object that they intended to intercept.

A key aspect of fixation selection in active tasks is the importance of the temporal allocation of gaze. This aspect is rarely emphasised in accounts (or indeed models) of

fixation behaviour when viewing static two-dimensional scenes. However, consistent relationships are found between the timing of gaze shifts and the timings of actions in many situations. Typically the eyes target an object about 0.5-1 second prior to manipulating it, and this timing is common across a wide variety of tasks including tea making (Land et al., 1999), driving (Land & Lee, 1994; Land & Tatler, 2001), music sight reading (Furneaux & Land, 1999), walking (Patla & Vickers, 2003), and reading aloud (Buswell, 1920). Moreover, successful completion of tasks may depend upon the correct temporal allocation of fixations: in cricket good and bad batsmen alike will look at the location on the crease where the ball bounces. The difference is that a good batsman will direct their eyes to this location about 100 ms before the ball arrives at the bounce point, whereas a poor batsman will direct their eyes to the same location at or just after the time that the ball arrives (Land & McLeod, 2000).

The correct spatiotemporal allocation of gaze in natural tasks requires that people must learn what to look at and when (Chapman & Underwood, 1998; Land, 2004; Land & Furneaux, 1997; Land & Tatler, 2001). Sailer, Flanagan and Johansson (2005) investigated how learning interacts with the spatiotemporal allocation of gaze in a visuomotor task. Their task required participants to guide a cursor to a series of targets on a monitor, controlled by a novel device with initially unknown mappings between actions and movements of the cursor. The task was initially very difficult but over a period of about 20 minutes participants became quite skilled at controlling the cursor. Of particular interest here is that the temporal relationship between gaze and the cursor changed dramatically over the learning period. Initially gaze lagged the cursor movements in time. However, by the time the participants were skilled at the task, gaze was allocated in an anticipatory manner. Moreover the timing was such that gaze led the cursor movements by around 0.4 seconds, which is in line with the typically observed lead by gaze over action that has been reported across a range of natural tasks. A similar

progression toward a greater lead time by the eyes over action can be found when comparing learner drivers to more experienced drivers (Land, 2006; Land & Tatler, 2001).

Learning can occur over a variety of timescales and can involve adapting behaviour in response to changes in the environment. For example, Jovancevic-Misic and Hayhoe (2009) showed that when walking toward other people what we learn about how someone is likely to behave when we encounter them is used to adapt our behaviour toward that person when we next encounter them. Oncoming pedestrians were assigned roles as potential colliders (who were asked to walk on collision courses toward the participant on each encounter) or avoiders (who were asked to avoid collision courses). Participants rapidly learnt who the potential colliders were and adapted their gaze behaviour such that they looked sooner and for longer at the potential colliders than at the avoiders. When the oncoming pedestrians switched roles, participants were able to adapt their responses after only a few encounters. Thus not only can gaze allocations be learnt 'on the fly' but also they can be adapted rapidly to changes in the environment.

If correct spatiotemporal allocation of gaze is central to skilled behaviour and this develops as we learn visuomotor skills, any model of gaze allocation in natural tasks should engage with this learning process.

Modelling eye movements in natural behaviour

At present, there is no overall model of gaze allocation in natural tasks. However, by identifying the key underlying principles of gaze selection in natural settings, it is possible to identify the aspects of eye movement behaviour that such a model should be able to explain (Tatler et al., 2011). As discussed above, it is clear that models of gaze allocation must engage with learning over multiple timescales. The reward sensitivity of the eye movement circuitry provides the neural underpinnings for reinforcement learning models of behaviour (Schultz, 2000; Montague & Hyman, 2004). Ballard, Hayhoe and

colleagues have developed models of natural behaviour based on the principles of reward (Sprague, Ballard, & Robinson, 2007; C. Rothkopf & Ballard, 2009; C. A. Rothkopf et al., 2007; Ballard & Hayhoe, 2009). In particular they have developed a model that guides a simulated walking agent through a virtual environment. This task contains three simultaneous sub-tasks that the agent must complete: staying within a defined path, avoiding certain obstacles and colliding with other obstacles. Each sub-task is associated with some reward value. For example, obtaining visual information that allows avoidance of an obstacle presumably provides secondary reward. The model assumes that information can only be gathered about one task at a time (much as the eyes can only be directed to a single location at a time) and that uncertainty will increase in the two unattended tasks. The decision to switch between sub-tasks is based on the uncertainty in the unattended tasks - that with the greatest uncertainty is attended next. Decisions about what to attend to are therefore made to maximise reward by reducing uncertainty that could result in sub-optimal actions. Framing the decision about where to look in terms of uncertainty reduction has been effective in explaining aspects of static scene viewing (Renninger, Verghese, & Coughlan, 2007; Najemnik & Geisler, 2005, 2008) as well as dynamic scene viewing. Such reward-based models are in their infancy but provide a compelling and promising direction for development in this field (Tatler et al., 2011).

Social factors in gaze selection

An often-neglected aspect of gaze control is the influence that the presence of another individual can have upon where we fixate. As we have seen, models of scene viewing typically focus on questions about low-level image properties or high-level task goals. But the mere presence of an individual in the scene can dramatically influence where we look. When presented with scenes containing people, observers preferentially fixate the faces of people in the scene (Birmingham, Bischof, & Kingstone, 2009).

Moreover, there is a strong tendency to orient gaze in the direction that another individual is looking (Driver et al., 1999; Friesen & Kingstone, 1998; Ricciardelli, Bricolo, Aglioti, & Chelazzi, 2002). In social scenes, participants spend more time looking at the object being fixated by a character in the scene than would be expected by chance (Fletcher-Watson, Findlay, Leekam, & Benson, 2008). When viewing sequences of photographs that told a story, participants were very likely to look at the actor's face and to saccade toward the object that was the focus of the actor's gaze direction (Castelhano, Wieth, & Henderson, 2007). When in a real environment, in which other people are present, we look far less at other people (or their eyes) than would be expected from lab-based studies of social attention (Gallup, Chong, & Couzin, 2012; Laidlaw, Foulsham, Kuhn, & Kingstone, 2011; Macdonald & Tatler, 2013), perhaps because to do so might signal a desire to engage in conversation with that individual: a situation that we often want to avoid. A particularly compelling situation in which people are strongly influenced by where another is looking is in the case of performance magic. One key component of performance magic is the misdirection of the audience. While there are a number of ways to achieve this, we have shown that the magician's gaze is a key component of misdirection in some performances (Kuhn & Tatler, 2005; Tatler & Kuhn, 2007). The effectiveness of this misdirection is greater during live performance (Kuhn & Tatler, 2005) than when watching a video of the performance (Kuhn, Tatler, Findlay, & Cole, 2008; Kuhn, Tatler, & Cole, 2009), reinforcing the importance of considering the setting when studying how our gaze allocation is influenced. Furthermore, the strong influence that the gaze direction of another individual has upon gaze allocation when viewing a scene underlines the need to consider this in models of eye movement behaviour.

Encoding information from the visual environment

Gaze allocation for visual sampling from the environment is only the first step in scene perception. I will now consider what is currently understood about the fate of the information that has been selected for sampling by the gaze control system. If the sampled information was all stored faithfully, then we might expect to find a close relationship between where we look and our subjective interpretation and experience of the scene. This possibility motivated some of the earliest work on eye movement behaviour, which considered the link between eye movements and the experience of illusions. Stratton found (much to his surprise) that there was no evidence that the experience or strength of illusion could be explained by eye movement patterns for the Muller-Lyer, Poggendorff or Zollner illusions (e.g. Stratton, 1906). In contrast, other contemporary researchers suggested that there may be evidence for links between where people look and the strength of experience of such illusions (C. H. Judd, 1905; Cameron & Steele, 1905).

The mapping between visual input and visual experience has underpinned a large volume of recent research. This is in part due to the inherent disconnect between the spatially restricted and temporally discontinuous sampling of the visual environment by the gaze system and the perceptually extensive and continuous experience we have of our surroundings. The obvious question to ask here is whether the continuous experience we have of our surroundings derives from an internal representation of our environment: stored information sampled from fixations could be used to construct internal representations, which could underpin a perceptual experience of the environment that is more extensive than that available from current visual input.

One thing that seems to be clear is that it is very unlikely that the representation takes the form of an integrated analogue representation of the visual information sampled in each fixation. Studies based on reading were the first to convincingly demonstrate that visual information may not be integrated from one fixation to the next. When reading

text with alternating letter cases, a global switch of all letter cases was not noticed by observers providing the switch occurred during a saccade (McConkie & Zola, 1979, Figure 7). Similarly, participants were unable to integrate two sets of lines that together made up a simple word if the views of the lines were separated by a saccade (O'Regan & Lévy-Schoen, 1983, Figure 7).

Change detection studies have since provided strong evidence of failures to integrate information across saccades when viewing scenes (Grimes, 1996; Rensink, O'Regan, & Clark, 1997). Changes go unnoticed when they are made to objects in scenes during brief interruptions to viewing such as blinks (Rensink, O'Regan, & Clark, 2000), saccades (Blackmore, Brelstaff, Nelson, & Troscianko, 1995) or flickers (Rensink et al., 1997). The initial interpretation of failures to detect changes in scenes was that this implied a failure to retain visually rich information beyond the end of a fixation (Rensink, 2002). Similar failure to retain visually rich information has also been suggested in the context of more natural, everyday settings (Tatler, 2001). Tatler (2001) found that if interrupted while making a cup of tea participants were able to report visually rich information about the locus of the interrupted fixation, but not about the locus of the preceding fixation. Not only did this imply a lack of retention of the content of previous fixations, but also the pattern of errors when asked to report the interrupted fixation content revealed insights into the fate of information once a saccade is executed. If the interruption occurred very soon after the start of the next fixation, the participants were likely to report the content of the penultimate rather than ultimate fixation. With increasing time into the new fixation, there was increasing probability of reporting the content of the new fixation. This result implies that pictorially rich information survives the end of a fixation and is retained until it is overwritten by the content of the new fixation soon after it begins (Figure 8).

Schemes of representation

There have been several different schemes of representation proposed that try to reconcile both our subjectively detailed visual experience and our inability to detect changes made to scenes during brief interruptions. One suggestion is that there is no internal representation of our surroundings of any kind (O'Regan & Noë, 2001). Under this interpretation, the high mobility of the eyes obviates the need for internal storage of information: if we need to know about a location in the environment we simply direct our eyes to that location. O'Regan and Noe (2001) suggested that our perceptions arise from the manner in which the information on the retina changes as we move our eyes, rather like earlier suggestions by Gibson (1979, 1950, 1966). Rensink (2000) favoured a less extreme position in which detailed representations are formed but are very selective and the detailed information survives only for as long as attention is focussed on a particular object. Rensink (2000) proposed that a limited number of proto-objects can be attended and bound together as an object representation, but once attention is disengaged from the proto-objects, the bound representation is also lost. In Rensink's scheme our internal representation is not limited to this bound object representation but is integrated with higher-level abstracted representations of the overall layout and gist of the scene. While both O'Regan and Rensink favour rather sparse accounts of representation, there is considerable evidence that what we retain and represent from each fixation may be considerably more detailed than was initially suggested by change detection studies.

A number of research groups have demonstrated that information is accumulated from scenes over time and across fixations (Hollingworth & Henderson, 2002; Irwin & Zelinsky, 2002; Melcher, 2006; Pertzov, Avidan, & Zohary, 2009; Tatler, Gilchrist, & Rusted, 2003). Irwin (Irwin, 1992; Irwin & Andrews, 1996) suggested that information about objects is accumulated in *object files*, which are temporary representations of the information pertaining to a range of properties of an object. Object files can be retained

for several seconds, but their number is limited (to around 3-5 object files), meaning that once all are full, any encoding of a new object is at the expense of an old object file. Hollingworth and colleagues (see Hollingworth, 2004, 2005, 2007) propose a more comprehensive and visually rich representation of the environment, which can survive over long timescales. Tatler, Gilchrist and Land (2005) found that the timescales and extents of information accumulation and retention were not unitary: different objects properties were encoded and retained in rather different ways and over different timescales. Tatler et al. (2005) found no evidence for encoding and retention of details of an object's shape or distance to neighbouring objects, but found that details of the object's colour, identity and position in the scene were encoded and retained. For the retained properties, patterns of encoding differed: identity and colour were encoded within a single fixation of the object, but position memory accumulated and improved over a number of fixations of the object. Divergence in timescales of representation was also found, with identity information being retained only transiently, whereas information about the colour and position of the object appeared less labile. These findings suggested that object representations may involve the independent encoding of a set of properties, encoded and retained over varying timescales.

Any representations of the environment are likely to influence ongoing viewing behaviour. Thus we can learn about representations from considering how they appear to influence ongoing behaviour such as saccade target selection. Saccades can be launched on the basis of remembered information (Karn, Møller, & Hayhoe, 1997), and brief previews of a scene alter subsequent search behaviour when the scene is inspected (Castelhano & Henderson, 2007). Oliva et al (2004) used panoramic scenes in which only some of the scene was visible at any time in order to consider the interplay between vision and memory in saccade planning. Participants forced to rely on either visual or remembered information alone were able to complete the search task. However, when both sources of information were present, search behaviour was dominated by the immediate visual

information. Taken together, these results argue that remembered information can influence ongoing gaze behaviour, but that for viewing static scenes gaze relocations are primarily under the control of immediate visual input.

Representation in active tasks

As explained in the first section of this chapter, many of our everyday settings and tasks are rather different from the typical picture-viewing paradigms that dominate the studies discussed above. One important departure is that we interact with objects in the environment rather than simply viewing them. Such interaction and manipulation of our environment may place very different demands on the representational system than simply looking at objects. Indeed, evidence from active tasks seems to paint a rather different picture of the likely nature of representations than the evidence discussed above.

When creating copies of models using coloured blocks, representations appear very sparse and limited in time. Ballard and colleagues (Ballard et al., 1992; Ballard, Hayhoe, & Pelz, 1995) showed that fixation strategies were less efficient than might be expected: for each cycle of selecting and placing a block there were two looks to the relevant block in the model (Figure 9). This result implied that each fixation of the block in the model was to extract a different property. The first was to extract the colour of the block so that a matching block could be selected from the source area. The second fixation was to encode the position of the block in the model for correct placement in the constructed copy. Over trials, the prevalence of this double-checking strategy declined, implying some build up of remembered information, but the continued observation of this strategy favours a rather sparse view of representation.

Triesch et al (2003) and Droll and Hayhoe (2007) used a virtual block-sorting task to consider the nature and stability of representations underlying visuomotor tasks. In both studies, blocks were sorted by different rules in different conditions, with each rule

emphasising different properties of the objects at different times in the task. Common to both studies was the inclusion of low-prevalence change trials in which a property of the object was changed during an eye movement, while the object was being manipulated. Triesch et al (2003) found that the likelihood of detecting a change to the object depended upon whether the features of the object were still relevant to the sorting task. When the change was only relevant to the rules for selecting an object, and not to where the block was placed, changes to the block were rarely detected (in 10% of trials). However, when the features of the object were relevant to both the selection and sorting decisions, a change to the object was detected in 45% of trials. This result implies that whether or not an object feature is retained (and hence available for change detection) depends upon whether it is still required for successful task completion. If the feature is no longer required it is no longer retained.

Droll and Hayhoe (2007) extended this finding by varying the participant's expectancy about the likely need for information later in the task. In one condition the same feature that was required for selecting and picking up a block was again required for the sorting task - thus it was entirely predictable that this information would be needed throughout the manipulation of the block. In this case, re-fixations of the block once picked up were rare, implying no need to re-encode information about the block. In a second condition, the feature required for selecting and picking up a block was predictable, but the feature required for sorting and placing the block was unpredictable and varied randomly. In this case it was not predictable that the information encoded for the selection decision would be needed again. In this unpredictable condition, re-fixations of the block during manipulation (between pickup and placing the block) were common, implying that resampling of the information was required in these cases. Importantly, frequent re-inspections of the object were found even when the sorting cue was the same as the selection cue, which occurred in 25% of trials due to the random selection of one of

the four defining features for the sorting rules. This result implies not only that representations are limited to what is required, but that participants only retain what they *expect* to need later in the task.

In many activities we engage in, we are required to move around in an extended environment. Behaviour in such an extended environment may again place rather different constraints on the representational system than are found when viewing images on a screen or conducting tasks in proximate space (Tatler & Land, 2011). A particular issue here is that of the reference frame in which representations should operate. In particular, there are a number of possible frames of reference in which to encode information about our surroundings, each with its own potential utility and limits for natural behaviour (Figure 10).

There has been considerable interest in the coordinate frame in which space may be represented in the brain (Andersen, Snyder, Bradley, & Xing, 1997; Burgess, 2006, 2008; Colby & Goldberg, 1999). It is clear that muscular movement plans must ultimately be coded in limb-centred coordinates. Similarly, visual information must initially be coded in retinotopic space. The parietal cortex appears to be equipped to deal with the interaction between a range of frames of reference, transforming between representations in different frames of reference (Chang, Papadimitriou, & Snyder, 2009). Recent accounts of the way we encode information about objects, places and routes in the world around us propose that we have two kinds of spatial representation: allocentric and egocentric (Burgess, 2006; Waller & Hodgson, 2006). The allocentric representation is map-like and indexed in world co-ordinates. In contrast the egocentric representation is based on directions relative to our current body position (Figure 11).

An appealing scheme for spatial representation in natural settings is to suggest that our on-line representation comprises the interplay between allocentric and egocentric representations of the surroundings (Tatler & Land, 2011, Figure 12). In our scheme the

on-line representation is fundamentally egocentric, containing low-resolution information about the identities and locations of objects throughout the 360 degree space around us. This representation therefore contains information from outside our current field of view, and which can be used to target movements of gaze or limbs irrespective of whether or not it is supplemented by direct visual information. Our view is that the allocentric representation is a longer-term representation of previously-viewed space which can be used to furnish the egocentric representation by a process similar to reading from a map. Thus our scheme suggests that moment-to-moment execution of gaze relocations and other behaviours is based upon the integration of direct visual input, the extended egocentric model and information read from enduring longer-term allocentric representations into the egocentric model. There is considerable evidence for the existence of both allocentric and egocentric representations in the brain, with the allocentric map located in the hippocampus and the medial temporal lobe, the egocentric model in the parietal lobe and translations from one to the other occurring in the retrosplenial cortex (Burgess, 2008).

One consequence of a scheme based around an egocentric on-line representation is that the representation must be constantly updated as we move around our environment, but such constant remapping of space can be conducted across saccades in LIP (Duhamel, Colby, & Goldberg, 1992).

A dual scheme of representation such as that which we have proposed offers an efficient coding scheme in which to plan our actions on the basis of a combination of immediate sensory input and remembered information. This scheme also allows differential reliance upon sensory and remembered information, with the potential to vary the relative reliance on these sources of information depending upon the availability and reliability of each: a flexibility which we know the gaze allocation system can exhibit (Brouwer & Knill, 2007). It is also interesting to speculate whether the egocentric model we describe might offer some bridge across the disconnect between disjointed sensory

input and smooth visual experience: the egocentric models provides the (albeit low-resolution) panoramic model that might provide enough detail to give the illusion of completeness that we experience in our visual interactions with the world.

What we retain from what we fixate is not only shaped by our expectations and task goals, but also by our physical interactions with objects. Specifically, we remember more about objects that we use than objects that we view but don't manipulate (Tatler et al., 2013). Moreover, manipulating objects confers benefits for later memory above and beyond those attributable to the relevance of objects to task goals (Tatler et al., 2013).

Conclusion

In this chapter we have considered what is currently understood about how we select information to sample from the environment and the subsequent fate of that information once the eyes are relocated to other locations in the scene.

In both cases we have seen that the setting in which these questions is studied can have a marked influence on the apparent mechanisms and processes that underlie these two aspects of scene perception. Much of our understanding of how we look at and remember scenes is derived from experimental paradigms using static photographic scenes. While how we look at images is an undeniably interesting and important question, it is equally important to consider the differences in findings between these situations and natural behaviour in real environments. For both gaze allocation and memory encoding there appear differences in the apparent underlying processes operating in real environments compared to those operating when viewing 2D static scenes.

What is clear from the material reviewed in this chapter is that there are similarities in the governing principles that influence both the spatiotemporal allocation of gaze and the encoding and retention of information from fixations. In both cases the task goals are central: we look at locations that offer information pertinent to completing the current

behavioural goal at the times when this information is required. Similarly, representations appear to be dependent upon what we require for a task and when we need it: if we are likely to need information again later in the task we retain it, whereas if we are not likely to need it again we do not retain a stored representation of the information.

Not only can we describe similar task-dependencies of information sampling and representation, but also we can see that both are based on what we expect to be important. Reward-based models of gaze allocation must be able to explain the anticipatory behaviour of the eye, typically being directed to places just before an action is carried out, or just before an event such as the arrival of a ball. As such, these schemes must be based upon the anticipated reward given our predictions about what is about to happen (Tatler et al 2011). A similar prominent role for prediction is seen in the stability of represented information. As Droll and Hayhoe (2007) elegantly demonstrated, whether or not information is retained in a block-sorting task depends upon the predictability of whether the information will be needed in the future. If it is not predictably of use later, then it is not retained. Thus it may well be that the traditionally separate field of eye movement control and scene memory share very similar substrates and governing principles.

References

Andersen, R., Snyder, L., Bradley, D., & Xing, J. (1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annual review of neuroscience*, *20*, 303–330.

Ballard, D. H., & Hayhoe, M. M. (2009). Modelling the role of task in the control of gaze. *Visual Cognition*, *17*(6-7), 1185–1204.

Ballard, D. H., Hayhoe, M. M., Li, F., & Whitehead, S. (1992). Hand-eye coordination during sequential tasks. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, *337*(1281), 331–338; discussion 338.

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory Representations in Natural Tasks. *Journal Of Cognitive Neuroscience*, *7*(1), 66–80.

Birmingham, E., Bischof, W., & Kingstone, A. (2009). Get real! Resolving the debate about equivalent social stimuli. *Visual Cognition*, *17*(6-7), 904–924.

Blackmore, S. J., Brelstaff, G., Nelson, K., & Troscianko, T. (1995). Is the Richness of Our Visual World an Illusion - Transsaccadic Memory for Complex Scenes. *Perception*, *24*(9), 1075–1081.

Borji, A., & Itti, L. (2013). State-of-the-Art in Visual Attention Modeling. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *35*(1), 185–207.

Brouwer, A.-M., & Knill, D. (2007). The role of memory in visually guided reaching. *Journal of Vision*, *7*(5), 1–12.

Burgess, N. (2006). Spatial memory: how egocentric and allocentric combine. *Trends in Cognitive Sciences*, *10*(12), 551–557.

Burgess, N. (2008). Spatial cognition and the brain. *Annals of the New York Academy of Sciences*, *1124*, 77–97.

Buswell, G. T. (1920). *An experimental study of the eye-voice span in reading.* Chicago: Chicago University Press.

Buswell, G. T. (1935). *How People Look at Pictures: A Study of the Psychology of Perception in Art.* Chicago: University of Chicago Press.

Bylinskii, Z., Judd, T., Borji, A., Itti, L., Durand, F., Oliva, A., & Torralba, A. (n.d.). *Mit saliency benchmark.*

Cameron, E. H., & Steele, W. M. (1905). The Poggendorff illusion. *Psychological Monographs*, *7*(1), 83–111.

Carmi, R., & Itti, L. (2006). Causal saliency effects during natural vision. In *Proceedings of the eye tracking research & application symposium, ETRA 2006, san diego, california, usa, march 27-29, 2006* (pp. 11–18).

Castelhano, M. S., & Henderson, J. M. (2007). Initial scene representations facilitate eye movement guidance in visual search. *Journal Of Experimental Psychology-Human Perception And Performance*, *33*(4), 753–763.

Castelhano, M. S., Wieth, M., & Henderson, J. M. (2007). I see what you see: Eye movements in real-world scenes are affected by perceived direction of gaze. *Attention in Cognitive Systems: Theories and Systems from an Interdisciplinary Viewpoint*, *4840*, 251–262.

Chang, S., Papadimitriou, C., & Snyder, L. H. (2009). Using a Compound Gain Field to Compute a Reach Plan. *Neuron*, *64*(5), 744–755.

Chapman, P., & Underwood, G. (1998). Visual search of driving situations: Danger and experience. *Perception*, *27*(8), 951–964.

Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annual review of neuroscience*, *22*, 319–349.

Cristino, F., & Baddeley, R. (2009). The nature of the visual representations involved in eye movements when walking down the street. *Visual Cognition*, *17*(6-7), 880–903.

Deubel, H., & Schneider, W. X. (1996, June). Saccade target selection and object recognition: evidence for a common attentional mechanism. *Vision Research*,

36(12), 1827–1837.

Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, *10*(10), 28: 1–17.

Driver, J., Davies, M., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers reflective visuospatial orienting. *Visual Cognition*, *6*(5), 509–540.

Droll, J. A., & Hayhoe, M. M. (2007). Trade-offs between gaze and working memory use. *Journal Of Experimental Psychology-Human Perception And Performance*, *33*(6), 1352–1365.

Duhamel, J. R., Colby, C. L., & Goldberg, M. E. (1992). The Updating of the Representation of Visual Space in Parietal Cortex by Intended Eye-Movements. *Science*, *255*(5040), 90–92.

Ehinger, K. A., Hidalgo-Sotelo, B., Torralba, A., & Oliva, A. (2009, August). Modeling Search for People in 900 Scenes: A combined source model of eye guidance. *Visual Cognition*, *17*(6-7), 945.

Einhauser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision*, *8*(14), 18: 1–26.

Epelboim, J. L., Steinman, R. M., Kowler, E., Edwards, M., Pizlo, Z., Erkelens, C. J., & Collewijn, H. (1995). The function of visual search and memory in sequential looking tasks. *Vision Research*, *35*(23-24), 3401–3422.

Epelboim, J. L., Steinman, R. M., Kowler, E., Pizlo, Z., Erkelens, C. J., & Collewijn, H. (1997). Gaze-shift dynamics in two kinds of sequential looking tasks. *Vision Research*, *37*(18), 2597–2607.

Erdmann, B., & Dodge, R. (1898). *sychologische Untersuchungen uber das Lesen auf experimenteller Grundlage*. Halle: Niemeyer.

Fecteau, J., & Munoz, D. (2006). Salience, relevance, and firing: a priority map for target selection. *Trends in Cognitive Sciences*, *10*(8), 382–390.

Fletcher-Watson, S., Findlay, J. M., Leekam, S. R., & Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception*, *37*(4), 571–583.

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, *8*(2), 6.1–17.

Friesen, C., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin and Review*, *5*(3), 490–495.

Frintrop, S., Rome, E., & Christensen, H. I. (n.d.). Computational visual attention systems and their cognitive foundations: A survey. *ACM Trans. on Applied Perception*, 2010.

Furneaux, S., & Land, M. F. (1999). The effects of skill on the eye-hand span during musical sight- reading. *Proceedings of the Royal Society of London Series B-Biological Sciences*, *266*(1436), 2435–2440.

Gallup, A. C., Chong, A., & Couzin, I. D. (2012). he directional flow of visual information transfer between pedestrians. *Biology Letters*, *8*(4), 520–522.

Gibson, J. J. (1950). *The Perception of the visual world* (1st ed.). Boston: Houghton Mifflin.

Gibson, J. J. (1966). *The Senses considered as Perceptual Systems.* New York: Appleton-Century-Crofts.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In K. Atkins (Ed.), *Perception: Vancouver studies in cognitive science* (pp. 89–110). New York: Oxford University Press.

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, *3*(1), 49–63.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*(11), 498–504.

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Chapter 25 - visual saliency does not account for eye movements during visual search in real-world scenes. In R. L. Hill, R. P. V. Gompel, M. H. Fischer, & W. S. Murray (Eds.), *Eye movements: A window on mind and brain* (p. 537 - 562). Oxford: Elsevier.

Hering, E. (1879). Über Muskelgeräusche des Auges. *Sitzungsberichte der Akademie der Wissenschaften in Wien. Mathematisch-naturwissenschaftliche Klasse. Abt. III*, *79*, 137–154.

Hollingworth, A. (2004). Constructing visual representations of natural scenes: The roles of short- and long-term visual memory. *Journal Of Experimental Psychology-Human Perception And Performance*, *30*(3), 519–537.

Hollingworth, A. (2005). The relationship between online visual representation of a scene and long-term scene memory. *Journal Of Experimental Psychology-Learning Memory And Cognition*, *31*(3), 396–411.

Hollingworth, A. (2007). Object-position binding in visual memory for natural scenes and object arrays. *Journal of Experimental Psychology-Human Perception and Performance*, *33*(1), 31–47.

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal Of Experimental Psychology-Human Perception And Performance*, *28*(1), 113–136.

Hong, B., & Brady, M. (2003). A topographic representation for mammogram segmentation. In *Medical image computing and computer-assisted intervention - miccai 2003, pt 2* (pp. 730–737). Univ Oxford, Med Vis Lab, Oxford, England.

Hooge, I., Over, E., Van Wezel, R., & Frens, M. A. (2005). Inhibition of return is not a foraging facilitator in saccadic search and free viewing. *Vision Research*, *45*(14), 1901–1908.

Irwin, D. E. (1992). Visual memory within and across fixations. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 146–165). New York: Springer-Verlag.

Irwin, D. E., & Andrews, R. (1996). Integration and accumulation of information across saccadic eye movements. In T. Inui & J. L. McClelland (Eds.), *Attention and performance xvi: Information integration in perception and communication* (pp. 125–155). Cambridge, MA: MIT Press.

Irwin, D. E., & Zelinsky, G. J. (2002). Eye movements and scene perception: Memory for things observed. *Perception & psychophysics*, *64*(6), 882–895.

Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition*, *12*(6), 1093–1123.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10-12), 1489–1506.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *20*(11), 1254-1259.

Jovancevic-Misic, J., & Hayhoe, M. (2009). Adaptive Gaze Control in Natural Environments. *Journal of Neuroscience*, *29*(19), 6234–6238.

Judd, C. H. (1905). The Müller-Lyer illusion. *Psychological Monographs*, *7*(1), 55–81.

Judd, T., Durand, F., & Torralba, A. (2012). *A benchmark of computational models of saliency to predict human fixations.*

Judd, T., Ehinger, K., Durand, F., & Torralba, A. (2009). Learning to predict where humans look. In *Ieee international conference on computer vision (iccv).*

Kanan, C., Tong, M., Zhang, L., & Cottrell, G. (2009). SUN: Top-down saliency using natural statistics. *Visual Cognition*, *17*(6-7), 979–1003.

Karn, K., Møller, P., & Hayhoe, M. M. (1997). Reference frames in saccadic targeting. *Experimental Brain Research*, *115*(2), 267–282.

Koch, C., & Ullman, S. (1985). Shifts in Selective Visual-Attention - Towards the Underlying Neural Circuitry. *Human Neurobiology*, *4*(4), 219–227.

Kuhn, G., & Tatler, B. W. (2005). Magic and fixation: now you don't see it, now you do. *Perception*, *34*(9), 1155–1161.

Kuhn, G., Tatler, B. W., & Cole, G. G. (2009). You look where I look! Effect of gaze cues on overt and covert attention in misdirection. *Visual Cognition*, *17*(6-7), 925–944.

Kuhn, G., Tatler, B. W., Findlay, J. M., & Cole, G. G. (2008). Misdirection in magic: Implications for the relationship between eye gaze and attention. *Visual Cognition*, *16*(2/3), 391–405.

Laidlaw, K. E., Foulsham, T., Kuhn, G., & Kingstone, A. (2011). Potential social interactions are important to social attention. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 5548–5553.

Land, M. F. (2004). The coordination of rotations of the eyes, head and trunk in saccadic turns produced in natural situations. *Experimental brain research*, *159*(2), 151–160.

Land, M. F. (2006). Eye movements and the control of actions in everyday life. *Progress in Retinal and Eye Research*, *25*(3), 296–324.

Land, M. F., & Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, *352*(1358), 1231–1239.

Land, M. F., & Horwood, J. (1995). Which parts of the road guide steering. *Nature*, *377*, 339–340.

Land, M. F., & Lee, D. N. (1994). Where we look when we steer. *Nature*, *369*(6483),

742–744.

Land, M. F., & McLeod, P. (2000). From eye movements to actions: how batsmen hit the ball. *Nature Neuroscience*, *3*(12), 1340–1345.

Land, M. F., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*(11), 1311–1328.

Land, M. F., & Tatler, B. W. (2001). Steering with the head: The visual strategy of a racing driver. *Current Biology*, *11*(15), 1215–1220.

Land, M. F., & Tatler, B. W. (2009). *Looking and acting: vision and eye movements in natural behaviour*. Oxford: OUP.

Macdonald, R. G., & Tatler, B. W. (2013). Do as eye say: Gaze cueing and language in a real-world social interaction. *Journal of Vision*, *13*(4), 6:1–12.

McConkie, G. W., & Zola, D. (1979). Is visual information integrated across successive fixations in reading? *Perception & psychophysics*, *25*(3), 221–224.

Melcher, D. (2006). Accumulation and persistence of memory for natural scenes. *Journal of Vision*, *6*(1), 8–17.

Montague, P., & Hyman, S. (2004). Computational roles for dopamine in behavioural control. *Nature*, *431*, 760–767.

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, *434*(7031), 387–391.

Najemnik, J., & Geisler, W. S. (2008). Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision*, *8*(3), 4:1–14.

Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research*, *45*(2), 205–231.

Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *Journal of vision*, *10*(8).

Nyström, M., & Holmqvist, K. (2008). Semantic override of low-level features in image

viewing-both initially and overall. *Journal of Eye Movement Research*, *2*(2), 2:1–11.

Oliva, A., Wolfe, J., & Arsenio, H. (2004). Panoramic search: The interaction of memory and vision in search through a familiar scene. *Journal Of Experimental Psychology-Human Perception And Performance*, *30*(6), 1132–1146.

O'Regan, J. K., & Lévy-Schoen, A. (1983). Integrating Visual Information from Successive Fixations - Does Trans-Saccadic Fusion Exist. *Vision Research*, *23*(8), 765–768.

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *The Behavioral and brain sciences*, *24*(5), 939–73; discussion 973–1031.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, *42*(1), 107–123.

Patla, A. E., & Vickers, J. N. (1997). Where and when do we look as we approach and step over an obstacle in the travel path? *Neuroreport*, *8*(17), 3661–3665.

Patla, A. E., & Vickers, J. N. (2003). How far ahead do we look when required to step on specific locations in the travel path during locomotion? *Experimental brain research*, *148*(1), 133–138.

Pelz, J. B., & Canosa, R. (2001). Oculomotor behavior and perceptual strategies in complex tasks. *Vision Research*, *41*(25-26), 3587–3596.

Pertzov, Y., Avidan, G., & Zohary, E. (2009). Accumulation of visual information across multiple fixations. *Journal of Vision*, *9*(10), 2.1–12.

Rayner, K. (1998). Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological bulletin*, *124*(3), 372–422.

Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network*, *10*(4), 341–350.

Renninger, L. W., Verghese, P., & Coughlan, J. (2007). Where to look next? Eye movements reduce local uncertainty. *Journal of Vision*, *7*(3), 6: 1–17.

Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition*, *7*(1-3), 17–42.

Rensink, R. A. (2002). Change detection. *Annual review of psychology*, *53*, 245–277.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*(5), 368–373.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (2000). On the failure to detect changes in scenes across brief interruptions. *Visual Cognition*, *7*(1-3), 127–145.

Ricciardelli, P., Bricolo, E., Aglioti, S. M., & Chelazzi, L. (2002). My eyes want to look where your eyes are looking: exploring the tendency to imitate another individual's gaze. *Neuroreport*, *13*(17), 2259–2264.

Rothkopf, C., & Ballard, D. H. (2009). Image statistics at the point of gaze during human navigation. *Visual neuroscience*, *26*(1), 81–92.

Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2007). Task and context determine where you look. *Journal of Vision*, *7*(14), 16.1–20.

Sailer, U., Flanagan, J. R., & Johansson, R. S. (2005). Eye-hand coordination during learning of a novel visuomotor task. *The Journal of neuroscience*, *25*(39), 8833–8842.

Schultz, W. (2000). Multiple reward signals in the brain. *Nature Reviews Neuroscience*, *1*(3), 199–207.

Siagian, C., & Itti, L. (2007). Biologically-inspired robotics vision monte-carlo localization in the outdoor environment. In *Ieee/rsj intelligent robots and systems* (pp. 1723–1730). San Diego, CA.

Smith, T., & Henderson, J. M. (2009). Facilitation of return during scene viewing. *Visual Cognition*, *17*(6-7), 1083–1108.

Sprague, N., Ballard, D. H., & Robinson, A. (2007). Modeling embodied visual behaviors. *ACM Transactions on Applied Perception*, *4*, 11.

Stainer, M. J., Scott-Brown, K. C., & Tatler, B. W. (2013). Looking for trouble: a description of oculomotor search strategies during live cctv operation. *Frontiers in Human Neuroscience*, *7*, 615. Retrieved from doi:10.3389/fnhum.2013.00615

Steinman, R. (2003). Gaze control under natural conditions. *The Visual Neurosciences*.

Stratton, G. M. (1906). Symmetry, linear illusions, and the movements of the eye. *Psychological Review*, *13*, 82–96.

Tatler, B. W. (2001). Characterising the visual buffer: real-world evidence for overwriting early in each fixation. *Perception*, *30*(8), 993–1006.

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, *7*(14), 4: 1–17.

Tatler, B. W. (Ed.). (2009). *Eye Guidance in Natural Scenes*. Hove, UK: Psychology Press.

Tatler, B. W., Baddeley, R., & Gilchrist, I. (2005). Visual correlates of fixation selection: effects of scale and time. *Vision Research*, *45*(5), 643–659.

Tatler, B. W., Gilchrist, I., & Land, M. (2005). Visual memory for objects in natural scenes: From fixations to object files. *Quarterly Journal of Experimental Psychology Section A-Human Experimental Psychology*, *58*(5), 931–960.

Tatler, B. W., Gilchrist, I. D., & Rusted, J. (2003). The time course of abstract visual representation. *Perception*, *32*(5), 579–592.

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, *11*(5), 5:1–23.

Tatler, B. W., Hirose, Y., Finnegan, S. K., Pievilainen, R., Kirtley, C., & Kennedy, A. (2013). Priorities for selection and representation in natural tasks. *Philosophical Transactions of the Royal Society B*, *368*, 20130066. Retrieved from http://dx.doi.org/10.1098/rstb.2013.0066

Tatler, B. W., & Kuhn, G. (2007). Don't look now: The magic of misdirection. In R. L. Hill, R. P. V. Gompel, M. H. Fischer, & W. S. Murray (Eds.), *Eye movements: A window on mind and brain* (pp. 697–714). Oxford: Elsevier.

Tatler, B. W., & Land, M. F. (2011). Vision and the representation of the surroundings in spatial memory. *Philosophical Transactions of the Royal Society B-Biological Sciences*, *366*(1564), 596–610.

Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, *2*(2), 5: 1–18.

t Hart, B., Vockeroth, J., Schumann, F., Bartl, K., Schneider, E., Konig, P., & Einhauser, W. (2009). Gaze allocation in natural stimuli: Comparing free exploration to head-fixed viewing conditions. *Visual Cognition*, *17*(6-7), 1132–1158.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, *113*(4), 766–786.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.

Triesch, J., Ballard, D., Hayhoe, M., & Sullivan, B. (2003). What you see is what you need. *Journal of Vision*, *3*(1), 86–94.

Wade, N. J., & Tatler, B. W. (2005). *The moving tablet of the eye: the origins of modern eye movement research*. Oxford: OUP.

Waller, D., & Hodgson, E. (2006). Transient and enduring spatial representations under disorientation and self-rotation. *Journal Of Experimental Psychology-Learning Memory And Cognition*, *32*(4), 867–882.

Wischnewski, M., Belardinelli, A., & Schneider, W. (2010). Where to look next? Combining static and dynamic proto-objects in a TVA-based model of visual attention. *Cognitive Computation*, *2*(4), 326–343.

Wischnewski, M., Steil, J., Kehrer, L., & Schneider, W. (2009). Integrating inhomogeneous processing and proto-object formation in a computational model of visual attention. *Human Centered Robot Systems*, 93–102.

Wolfe, J. (2007). Guided Search 4.0: Current Progress with a model of visual search. In W. Gray (Ed.), *Integrated models of cognitive systems* (pp. 99–119). New York: OUP.

Xu, T., Kuehnlenz, K., & Buss, M. (2010). Autonomous Behavior-Based Switched Top-Down and Bottom-Up Visual Attention for Mobile Robots. *Ieee Transactions on Robotics*, *26*(5), 947–954.

Yarbus, A. L. (1967). *Eye movements and vision.* New York: Plenum Press.

Zelinsky, G. J. (2008). A Theory of Eye Movements During Target Acquisition. *Psychological Review*, *115*(4), 787–835.

Author Note

Comments may be sent to the author at b.w.tatler@dundee.ac.uk

Figure Captions

Figure 1. Left, eye movement recording of one participant viewing Hokusai's *The Wave*. Right, *The Wave* by Hokusai.

Figure 2. Schematic of Itti and Koch's (2000) salience model, redrawn for Land and Tatler (2009)

Figure 3. Left, eye movements of an individual viewing the Chicago Tribune Tower with no specific instructions. Right, eye movements of the same individual when instructed to look for a face at a window in the tower.

Figure 4. Recordings of one participant viewing *The Unexpected Visitor* seven times, each with different instructions prior to viewing. Each record shows eye movements collected during a 3-minute recording session. The instructions given were (a) Free examination. (b) Estimate the material circumstances of the family in the picture. (c) Give the ages of the people. (d) Surmise what the family had been doing before the arrival of the unexpected visitor. (e) Remember the clothes worn by the people. (f) Remember the position of the people and objects in the room. (g) Estimate how long the unexpected visitor had been away from the family. (Illustration adapted from Yarbus, 1967, Figure 109, for Land and Tatler, 2009)

Figure 5. Top Left, eye movements of 40 subjects during the first second of viewing *The Wave*. Top Right, eye movements of 40 subjects during the final second of viewing *The Wave*. Bottom Left, fixation locations for 14 observers during the first second of viewing a photographic scene (from Tatler et al., 2005). Bottom right, data from the same 14 participants recorded during the 5th second of viewing the same photographic scene.

Figure 6. The central bias in scene viewing. Strong central biases in fixation distributions

(bottom row) are found for scenes irrespective of their feature distributions (middle row). Data from Tatler (2007).

Figure 7. Participants read text of alternating uppercase and lowercase letters as shown in the upper line. When the eye (black circles denote fixations, the arrows saccades) passed an invisible boundary, shown by the dashed line, the case of every letter in the display was changed so that by the time the eye landed for the next fixation the text was as shown in the lower line. Participants did not notice the change and there were no measurable differences in fixation duration or saccade amplitude around the time of the change. Redrawn from McConkie and Zola (1979). (b) Participants fixated a central marker until a peripheral target appeared. When the target appeared an array of lines also appeared between the centre of the screen and the peripheral target. When the participant launched an eye movement toward the target, the lines changed to a different array. The lines were meaningless alone, but if the pre- and post-saccade lines were fused they would form one of three French words. Participants were incapable of reporting these words. Redrawn from O'Regan and Levy-Shoen (2003) for Land and Tatler (2009).

Figure 8. Schematic of Tatler's proposed transient retention of visually rich information across saccades, until overwritten by the content of the new fixation.

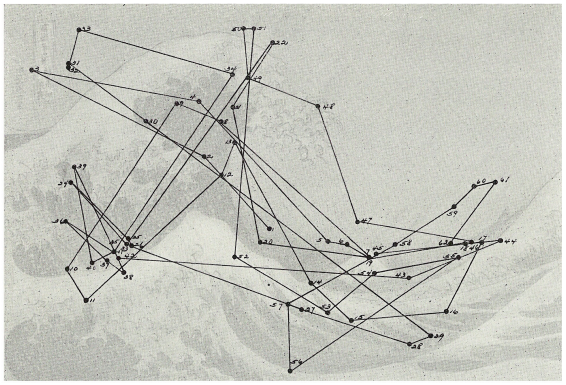
Figure 9. Ballard's block-copying task, illustrating the most common visual strategy by participants. Typically, participants fixate a block in the model (1) before fixating a block of the corresponding colour in the source area (2). Once the block is picked up and in transit towards the copy area (dashed grey arrow), a refixation of the block in the model is made (3), presumably to gather information about where to place the selected block. Finally, the location at which the block will be placed is fixated (4).

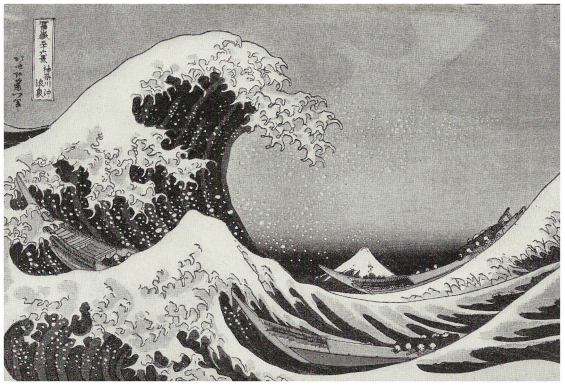
Figure 10. Frames of reference for visuomotor tasks. The required movement to grasp the

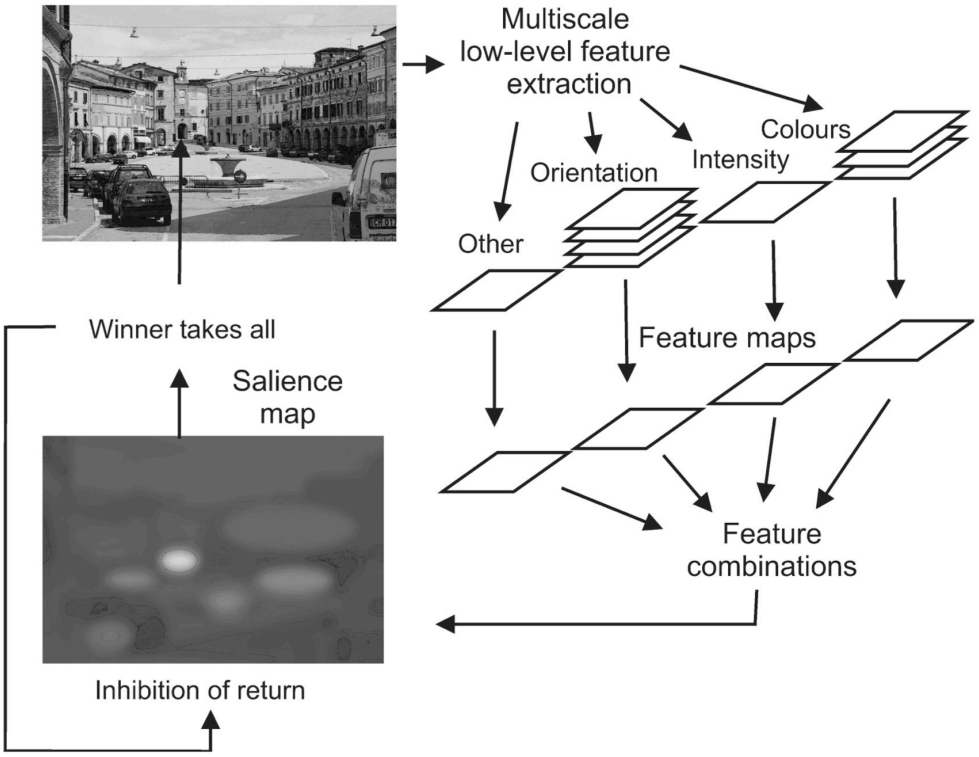
mug is the angle from arm to target. This is the angle from body-to-arm minus the sum of the angles from target-to-fovea, eye-in-head and head-on-body. In practice, eye, head and body are often aligned before such a grasp movement, but such alignment is not essential.

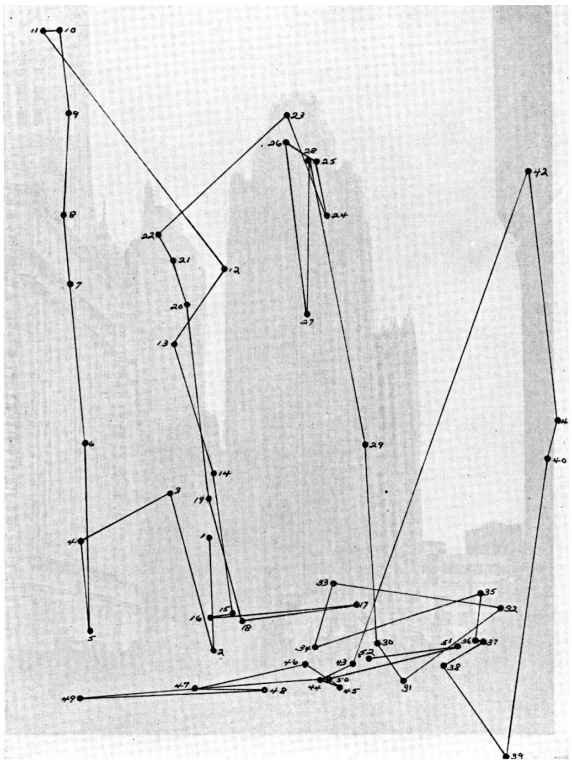
Figure 11. (a) Allocentric representation of a kitchen. This is independent of location and viewpoint. (b) Egocentric representation showing that the action required to reach the mug depends on the relation of the mug to the actor in egocentric space.

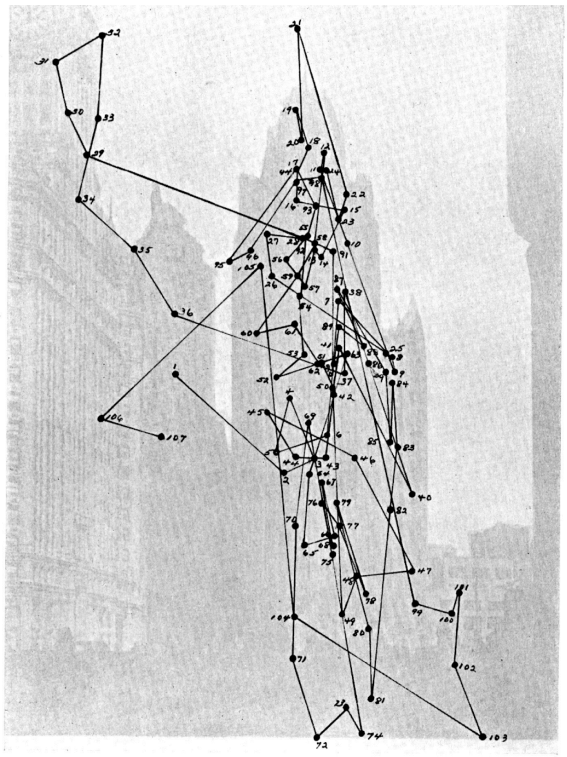
Figure 12. Planning to locate and reach for a target (T). (a) The interplay between vision (oval centred around the fovea, F) and egocentric representation (grey background centred around the head). In this example, we consider the situation where the observer intends to reach for a target (T) that is outside the field of view and not currently foveated. First, a gaze shift is planned to bring the fovea to bear upon the target. This gaze shift is planned using information from the egocentric model, which itself is furnished by information from ambient vision in the past and from the allocentric representation. (b) The situation after the gaze shift to the new target (T). As gaze shifts clockwise from F to T, so vision is re-centred around T and the egocentric map in the head is rotated anti-clockwise to re-centre around T. The manual reach can now be executed using motor commands planned using information provided by the fovea.

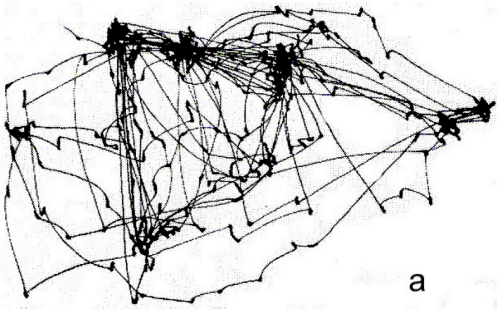


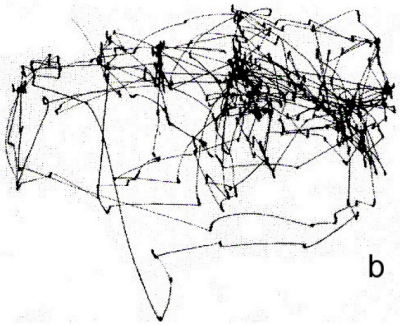


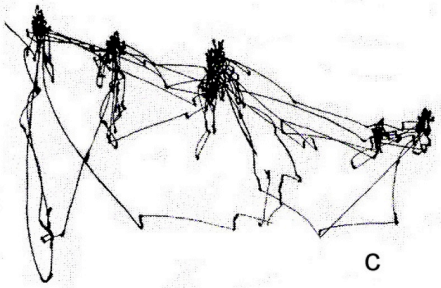

Multiscale
low-level feature
extraction
Colours
Intensity
Orientation
Other
Feature maps
Feature
combinations
Winner takes all
Salience
map
Inhibition of return

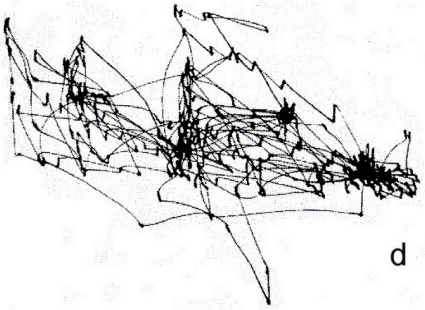


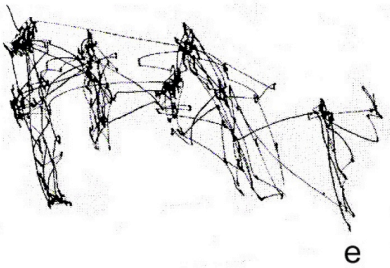


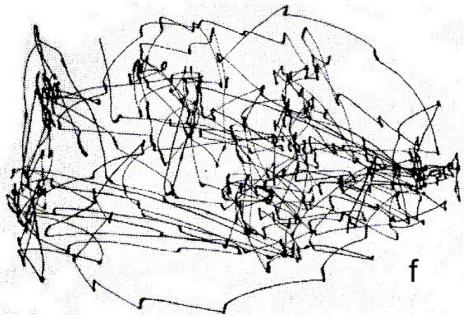

a

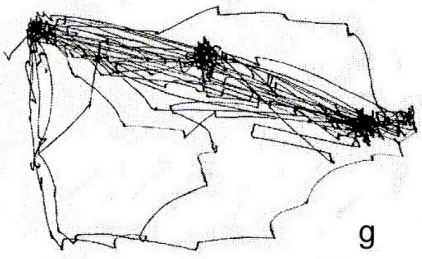

b

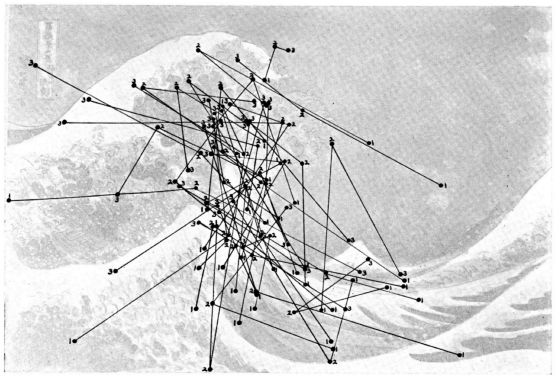

c

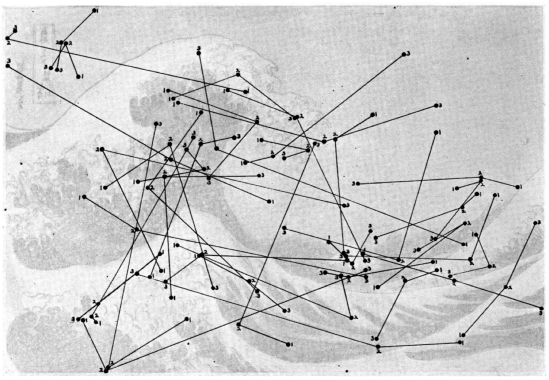

d

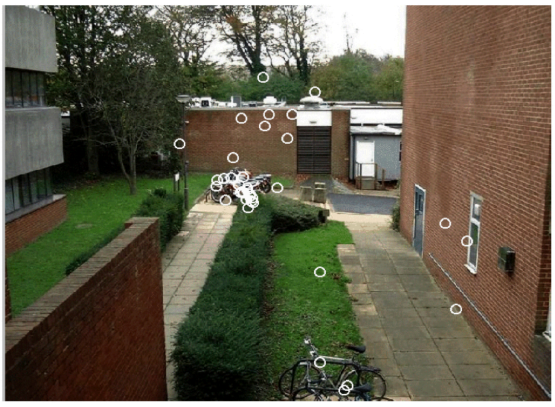

e

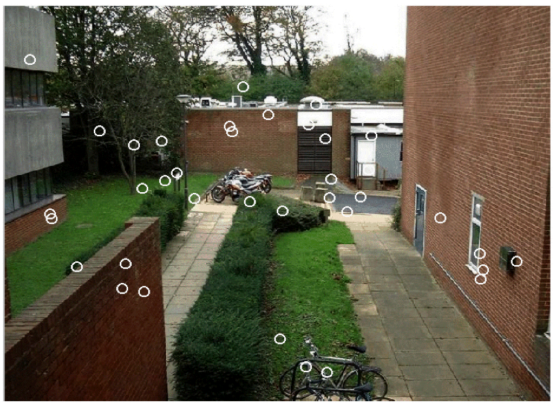

f

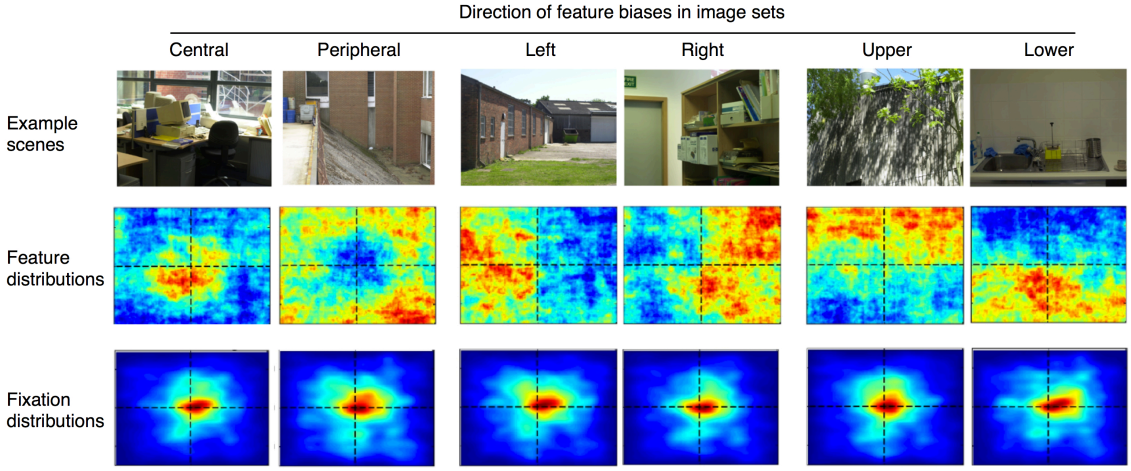

g

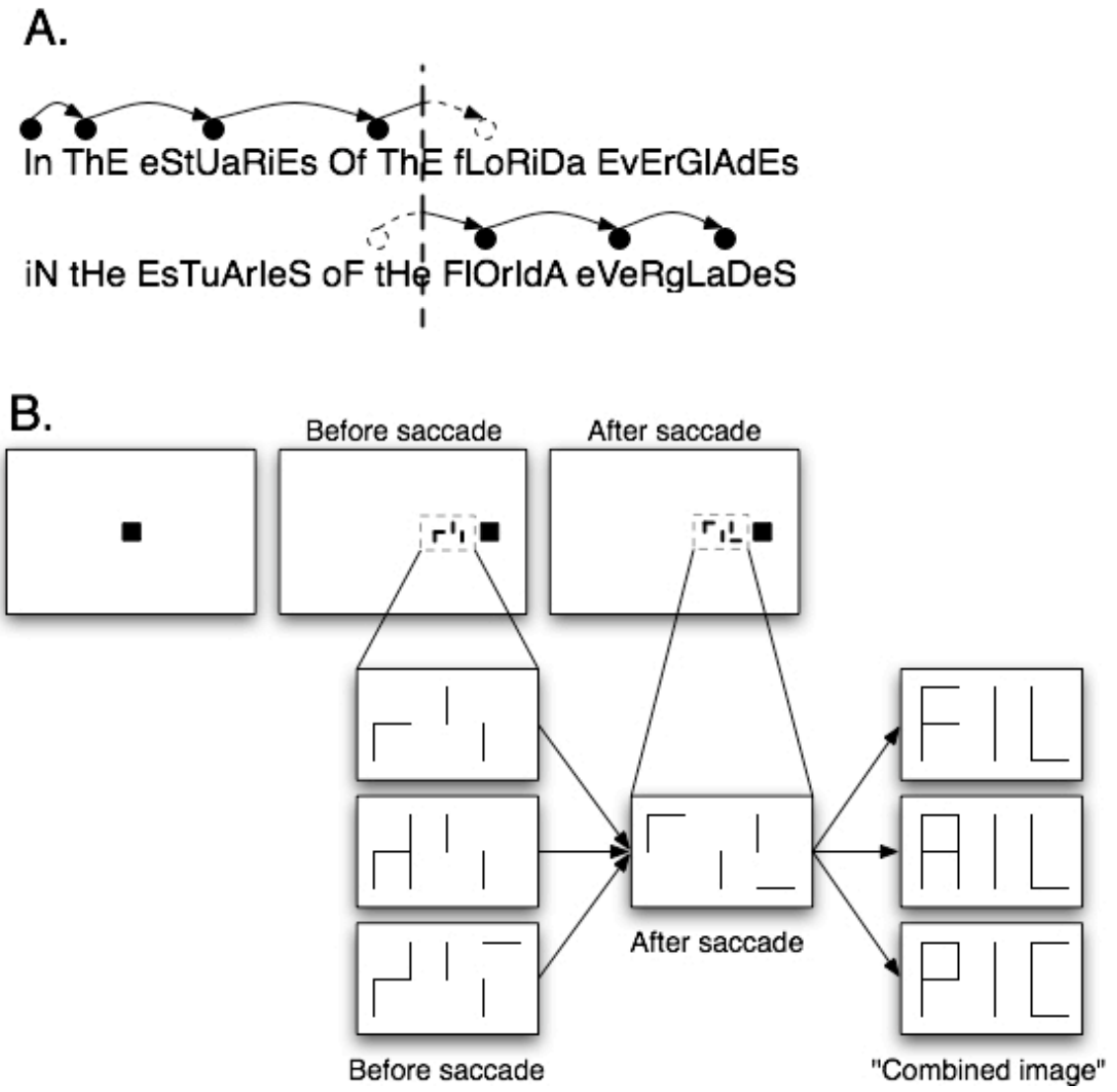


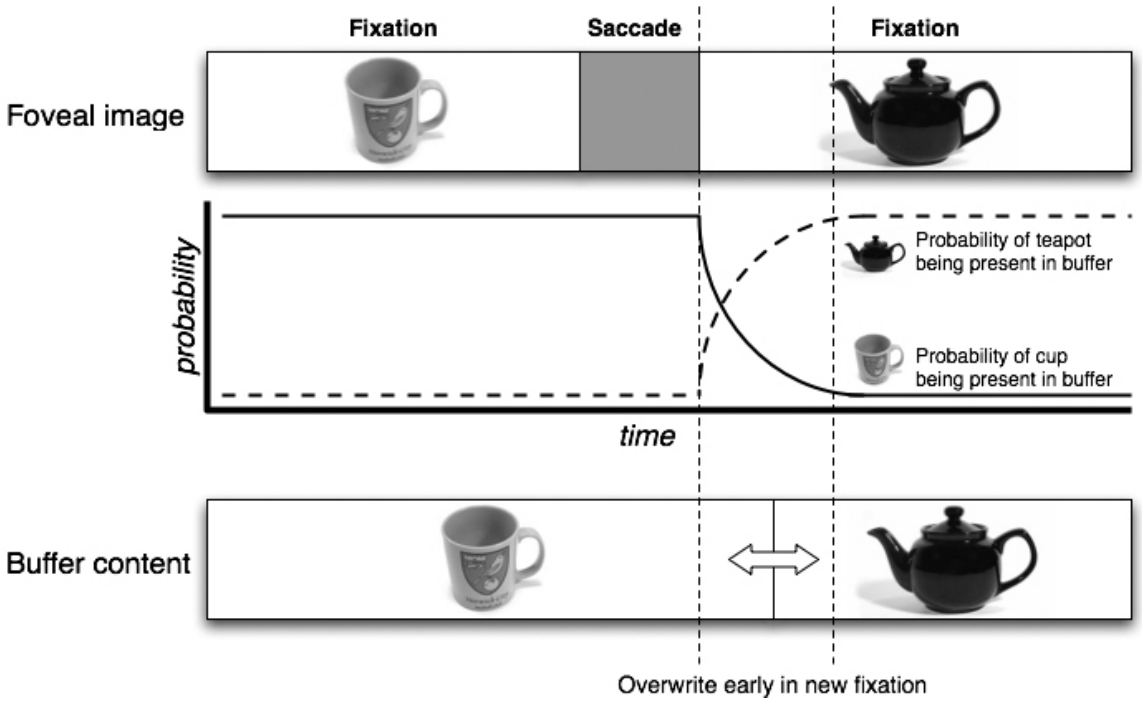







Direction of feature biases in image sets
Central
Peripheral
Left
Right
Upper
Lower
Example scenes
Feature distributions
Fixation distributions


A.
In ThE eStUaRiEs Of ThE fLoRiDa EvErGlAdEs
iN tHe EsTuArIeS oF tHe FlOrIdA eVeRgLaDeS
B.
Before saccade
After saccade
After saccade
Before saccade
"Combined image"


Fixation
Saccade
Fixation
Foveal image
probability
Probability of teapot being present in buffer
Probability of cup being present in buffer
time
Buffer content
Overwrite early in new fixation